

Fusión de Respuestas en la Búsqueda de Respuestas Multilingüe

Rita M. Aceves-Pérez, Manuel Montes-y-Gómez, Luis Villaseñor-Pineda

Laboratorio de Tecnologías del Lenguaje, Depto. Ciencias Computacionales,
Instituto Nacional de Astrofísica, Óptica y Electrónica, México.
{rmaceves, mmontesg, villasen}@inaoep.mx

Resumen: El presente trabajo describe los principales retos de los sistemas de Búsqueda de Respuestas (BR) Multilingües. Concretamente se aborda el problema de la combinación o fusión de respuestas candidatas provenientes de búsquedas en colecciones en diferentes idiomas. Este trabajo compara diferentes técnicas de fusión, propuestas en otras áreas de tratamiento multilingüe, a la problemática específica de la BR. A partir del análisis de los resultados se discuten las características esenciales que deberá considerar un método de fusión orientado a la BR multilingüe.

Palabras clave: Búsqueda de Respuestas Multilingüe, Fusión de respuestas, CLEF.

Abstract: This paper describes the main challenges of multilingual question answering. In particular, it focuses on the problem of answer fusion, which concerns the combination of answers obtained from collections in different languages. Basically, it compares a set of techniques for answer fusion, previously used in other multilingual tasks, in the problem of multilingual question answering. On the basis of the achieved results, we discuss the essential characteristics for an answer fusion method to be especially suited to the task of multilingual question answering.

Keywords: Multilingual Question Answering, Answer Fusion, CLEF.

1 Introducción

Los sistemas de búsqueda de respuestas (BR) son aquellos que responden de manera concisa a preguntas concretas formuladas en lenguaje natural. Uno de los principales retos que afrontan en la actualidad estos sistemas es el multilingüismo, situación en la cual se desea que el sistema de BR responda a preguntas formuladas en un idioma diferente al de la colección de búsqueda. Bajo tal situación es posible distinguir dos tipos de sistemas de BR: los sistemas translingües y los multilingües. Un sistema translingüe es aquel donde la colección de búsqueda está en un idioma diferente al de la pregunta. Y un sistema multilingüe realiza la búsqueda de la respuesta en dos o más colecciones, todas ellas en idiomas diferentes.

Cada uno de estos sistemas involucra diferentes retos. El presente trabajo aborda uno de los desafíos que los sistemas BR multilingües deben resolver.

Un sistema de BR multilingüe puede ser visto como un conjunto de sistemas de BR monolingües trabajando de manera coordinada, donde cada sistema se dedica a la extracción de respuestas bajo una colección de búsqueda en particular. Por supuesto, dos pasos más son de vital importancia: la traducción de la pregunta a los diferentes idiomas de las colecciones de búsqueda; y la combinación (o fusión) de las respuestas entregadas por cada sistema de BR monolingüe. La figura 1 muestra un esquema de un sistema de BR multilingüe.

Durante la traducción de la pregunta a diferentes idiomas el principal problema es disminuir el impacto negativo de la traducción

automática. Una solución a este problema consiste en emplear 2 o más traductores para posteriormente, a partir de diferentes criterios, elegir la mejor traducción, o crear una nueva reformulación. Detalles de estos métodos pueden consultarse en (Aceves-Pérez *et al.*, 07)

Con las n traducciones, tantas como colecciones de búsqueda existan, se realiza cada una de las búsquedas monolingües correspondientes.

Finalmente, con el conjunto de listas ordenadas de respuestas candidatas, provenientes de los sistemas de BR monolingües, se inicia el paso de fusión. El objetivo de este último paso, es la integración de las respuestas en una única lista ordenada. Esta integración deberá considerar tanto el orden de las respuestas en las listas, así como su repetida aparición en ellas. El objetivo central de este trabajo es el análisis del comportamiento de las técnicas de fusión, propuestas en otras áreas de tratamiento multilingüe, a la problemática específica de la BR. Para ello, se han aplicado algunas de las técnicas de fusión básicas y se han analizado y comparado sus resultados.

El resto del artículo está organizado de la siguiente manera. La sección 2 presenta el trabajo relacionado tanto en sistemas de BR translingües y multilingües, así como en las técnicas de fusión propuestas en la Recuperación de Información Multilingüe. La sección 3 presenta brevemente las estrategias más prometedoras para la fusión de listas de respuestas. La sección 4 describe los experimentos realizados y analiza los resultados alcanzados. Finalmente, la sección 5 presenta las conclusiones de este trabajo.

2 Trabajos relacionados

Dado el gran interés en el desarrollo de los sistemas de BR, a partir del año 2003 el *Cross-Language Evaluation Forum* (CLEF) inicia la evaluación de sistemas monolingües y translingües de BR en idiomas europeos (Magnini *et al.*, 03).

En la actualidad podemos encontrar muchos ejemplos de sistemas de BR translingües. La arquitectura clásica de estos sistemas consta de dos componentes: un módulo de traducción de la pregunta y un sistema de BR monolingüe. Su

funcionamiento consiste en: (i) la pregunta – formulada en un lenguaje fuente– es traducida al lenguaje de la colección de búsqueda –el lenguaje destino; y (ii) la traducción se entrega a un sistema de BR monolingüe afín al lenguaje destino. Dada la naturaleza de las preguntas, comúnmente no se realiza la traducción de la respuesta al lenguaje fuente.

Los primeros sistemas translingües implementaban el módulo de traducción utilizando directamente un traductor automático (Echihabi *et al.*, 03; Jijkoun *et al.*, 04; Perret, 04). En estos primeros sistemas se pudo notar que el impacto del error en la traducción era considerable. Para disminuir este impacto se propusieron métodos alternos. Un primer enfoque consiste en asegurar la correcta traducción de los términos claves de la pregunta. Es el caso del trabajo de (Tanev *et al.*, 2004) donde, apoyándose en diccionarios bilingües y EuroWordNet, genera su propia traducción. Otro enfoque, consiste en utilizar dos o más traductores. Esta idea se basa en la suposición de que el acuerdo entre ellos tendrá mayores probabilidades de ser correcto. Por ejemplo, el método de (Laurent *et al.*, 05) realiza una traducción triangulada usando al inglés como idioma pivote. Otro ejemplo es el trabajo de (Sutcliffe *et al.*, 05), el cual genera una traducción término-por-término combinado dos traductores diferentes y un diccionario. Finalmente, también encontramos otros trabajos que buscan medios pertinentes para resaltar el acuerdo entre los traductores. Por ejemplo, (Neuman & Sacaleanu, 05) construyen una bolsa de palabras expandida usando los términos de diferentes traducciones de la pregunta, así como sus sinónimos extraídos de EuroWordNet. Otro ejemplo, es el trabajo de (Aceves-Pérez, *et al.* 07) el cual reconstruye una nueva traducción a partir de las secuencias comunes entre varios traductores.

Respecto a los sistemas de BR multilingüe existe, en nuestro conocimiento, únicamente el trabajo de (García-Cumbreras *et al.*, 06). En éste se propone un sistema de BR basado en un subsistema de Recuperación de Información Multilingüe. Sin embargo, en la parte experimental sólo se presentan resultados para el caso monolingüe del español.

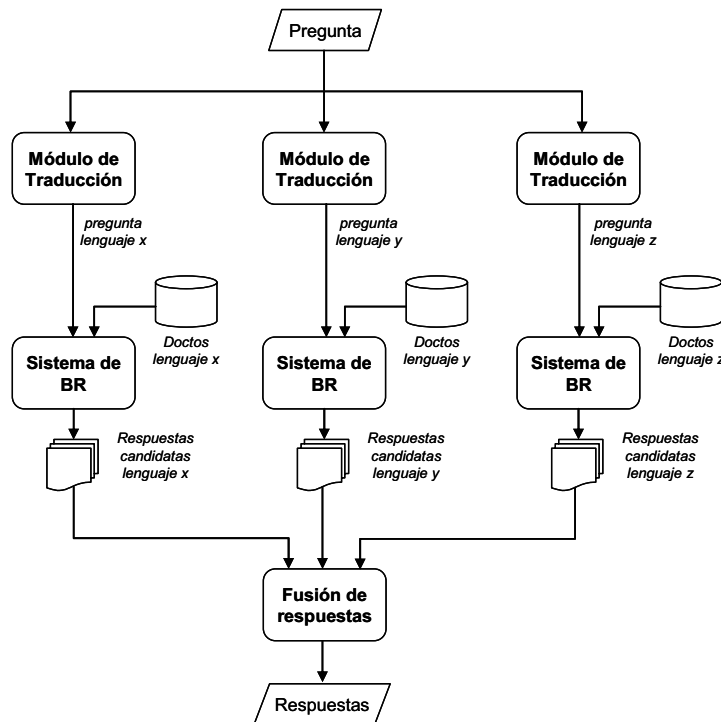


Figura 1. Esquema general de un sistema de BR multilingüe

Por otro lado, el problema de la fusión de listas de respuestas –provenientes de diferentes colecciones en diferentes idiomas– no ha sido abordado en el contexto de Búsqueda de Respuestas. Los trabajos más cercanos realizan búsquedas en más de una colección, siempre en el mismo idioma, y aplican diferentes técnicas para combinar (o fusionar) las respuestas provenientes de cada colección. Por ejemplo, los trabajos de (Echihabi *et al.*, 03, Jijkoun *et al.*, 04) realizan búsquedas en secuencia sobre las diferentes colecciones, donde la lista de respuestas calculada en el paso previo es confirmada (o filtrada) por la búsqueda subsiguiente. (Chu-Carroll, 03) también realiza búsquedas en varias colecciones, todas ellas en inglés, una vez que obtiene las 5 respuestas candidatas con mayor calificación de cada una de las colecciones, las mezcla y reordena de acuerdo a su calificación, si existe alguna respuesta en más de una colección sus calificaciones se suman (véase la estrategia *CombSum* en la siguiente sección). En (Sangoi y Mollá, 04) la búsqueda de respuestas se realiza sobre la Web recuperando extractos con diferentes motores de búsqueda. Estos extractos son pesados de acuerdo a la eficiencia del motor de búsqueda. Posteriormente se combinan, y realizando un cálculo estadístico sobre las entidades nombradas, se reordenan.

Una conclusión que podemos extraer de estos métodos es que la búsqueda en diferentes colecciones permite mejorar el comportamiento de los sistemas de BR.

3 Fusión de respuestas candidatas

La idea central de este trabajo consiste en evaluar algunos de los métodos tradicionales utilizados por los sistemas de Recuperación de Información Multilingüe (CLIR por sus siglas en inglés) al colocarlos en el contexto de la BR multilingüe. Cabe señalar que no se trata de problemas idénticos, la problemática de la fusión de listas de documentos en CLIR tiene por objetivo crear una nueva lista de documentos donde los más relevantes, sin importar su idioma, estén en las primeras posiciones de la lista. En la BR multilingüe el orden de las primeras posiciones es en extremo importante. Existe una gran diferencia dar la respuesta correcta en la primera posición, a darla en la segunda o quinta posición. A continuación se describen brevemente las estrategias de fusión tradicionales propuestas dentro de la CLIR.

RoundRobin. Esta estrategia toma la respuesta de más alta calificación de cada una de las listas y las coloca alternadamente en la nueva lista. Posteriormente, toma las segundas mejor

calificadas y las coloca en la nueva lista. Esta operación se repite sucesivamente hasta terminar las listas. En este caso, las respuestas duplicadas son descartadas.

RSV (Raw Score Value). Esta estrategia toma en cuenta las calificaciones de cada respuesta determinadas en la fase anterior. Las listas de respuestas se mezclan en una única lista y reordenándose de acuerdo a dicha calificación. Cuando una respuesta aparece en más de una lista, se suman sus calificaciones.

CombSum. Esta técnica de fusión, propuesta por (Lee, 97), puede considerarse como una combinación entre *RoundRobin* y *RSV*. Este método asigna una calificación de $21-i$ a las 21 primeras respuestas de cada una de las listas – ordenadas descendientemente– siendo i la posición de la respuesta. Cualquier respuesta después de la posición 21 se le asigna una calificación de 0. De esta forma, la primera respuesta (en todas las listas) queda con una calificación de 20, la segunda con 19 y así sucesivamente. Finalmente, las listas se mezclan y se reordenan atendiendo a la nueva calificación. En caso de que una respuesta se encuentre en más de una lista sus calificaciones se suman.

CombMNZ. Esta técnica, también propuesta por (Lee, 97), es una variante de la anterior. En este caso se asigna a cada respuesta una calificación tal como se describe en *CombSum*, pero esta calificación es multiplicada por el número de colecciones en la que se encontró dicha respuesta.

Como puede observarse estas estrategias aprovechan de diferente forma la información de las listas. En el caso de *RoundRobin*, la estrategia más simple, se da prioridad al orden relativo en las listas y no se aprovecha la aparición repetida de una respuesta. Por otro lado, esta estrategia también es sensible al orden en que se procesan las listas. Dando los mejores lugares en la lista final a las respuestas de la primera lista de respuestas procesada. Las otras tres estrategias revisadas abordan estos inconvenientes buscando esquemas de pesado que mejoren el ordenamiento final.

4 Experimentos y Resultados

En esta sección se muestran los resultados al evaluar las estrategias de fusión mencionadas en la sección anterior. A continuación se describen los diferentes recursos usados en los experimentos.

4.1 Recursos

Las colecciones de búsqueda. Se usaron 3 colecciones de búsqueda en tres idiomas diferentes, todas ellas de notas periodísticas. La colección en español con 454,045 documentos (aprox. 1Gb.), la de francés con 129,806 documentos (325 Mb), y la de italiano con 157,558 documentos (350 Mb). Las colecciones corresponden a las utilizadas en el CLEF 2005.

El conjunto de pregunta y respuestas. Para la selección de las preguntas y sus respuestas se tomó como base el corpus Multi-eight del CLEF. De este corpus se extrajeron preguntas para cada uno de los tres idiomas. Se tuvo especial cuidado en seleccionar preguntas cuya respuesta estaba en las listas de respuestas otorgadas por los sistemas de BR monolingües. De esta forma se creó un conjunto de 170 preguntas. Para poder evaluar el alcance de las estrategias de fusión se identificó el subconjunto de preguntas que tienen respuesta en una sola colección de búsqueda, así como el subconjunto de preguntas cuya respuesta está presente en más de una colección de búsqueda. Como es de suponer, es precisamente en este segundo subconjunto de preguntas que las estrategias de fusión tendrán un mayor impacto. La tabla 1 muestra la distribución de preguntas contestadas por colección de búsqueda.

El sistema de BR monolingüe. El sistema de BR utilizado fue el sistema TOVA (Montes-y-Gómez *et al.*, 05), un sistema basado únicamente en información léxica, lo que lo hace prácticamente independiente del idioma. Este sistema obtuvo la mejor posición en el ejercicio monolingüe en italiano, y la segunda mejor posición en el ejercicio monolingüe en español del CLEF 2005.

<i>Colección</i>	<i>No. Preguntas</i>
Español	37 (21%)
Francés	21 (12%)
Italiano	15 (9%)
Español-Francés	20 (12%)
Español-Italiano	25 (15%)
Francés-Italiano	23 (14%)
Español-Francés-Italiano	29 (17%)

Tabla 1. Distribución de preguntas en función de la(s) colección(es) donde se encuentran sus respuestas.

4.2 Evaluación

El criterio de evaluación de relevancia de las respuestas es el mismo que el propuesto en el CLEF. Dado que el objetivo de nuestro sistema de BR multilingüe es mejorar la precisión, se tomó como punto de referencia (*baseline*) el comportamiento del sistema de BR en su versión monolingüe para el español.

La precisión es la cantidad de preguntas respondidas correctamente en relación a la cantidad total de preguntas. Para observar con mayor claridad el comportamiento de las diferentes estrategias se midió la precisión a una, tres y cinco posiciones. Para el caso de tres o cinco posiciones, la pregunta se considera correctamente respondida si la respuesta apropiada se encuentra entre las primeras tres o cinco respuestas candidatas. Cabe resaltar que el proceso de evaluación se realizó de manera manual evitando errores debido a la necesidad de un apareo perfecto.

4.3 Experimentos

Los experimentos realizados están orientados a evaluar el comportamiento de las técnicas de fusión sobre diferentes listas de respuestas candidatas en diferentes idiomas.

El primer paso consistió en obtener las listas de respuestas candidatas. Para ello se realizaron tres corridas monolingües: una en español, otra en francés y una última en italiano. Cabe señalar que las preguntas usadas en las 3 corridas del sistema de BR fueron las mismas pero en diferentes idiomas. Las preguntas se tomaron directamente del corpus para evitar el ruido por la traducción automática de las preguntas.

El siguiente paso consistió en traducir las listas de respuestas en francés e italiano al

español, en este caso si se utilizó un traductor automático (Systran). Finalmente, se han aplicado las diferentes técnicas de fusión a las tres listas.

En la tabla 2 se muestran los resultados obtenidos con cada estrategia de fusión, distinguiendo la precisión calculada a la primera, a la tercera y a la quinta posición. A manera de referencia se incluyó la precisión obtenida en el ejercicio monolingüe en español (última fila).

<i>Método</i>	<i>1a.Pos.</i>	<i>3a.Pos.</i>	<i>5a.Pos.</i>
RSV	0.44	0.61	0.69
RoundRobin	0.45	0.68	0.74
CombSum	0.42	0.66	0.75
CombMNZ	0.42	0.62	0.70
<i>Monolingüe</i>	<i>0.45</i>	<i>0.57</i>	<i>0.64</i>

Tabla 2. Precisión con los diferentes métodos de fusión de respuestas.

Como puede observarse en la tabla 2, los mejores resultados se alcanzaron con el método de *RoundRobin*, incluso superando los resultados del sistema monolingüe con precisión a 3 y 5 posiciones. Sin embargo, este resultado fue inesperado, ya que este método no considera la redundancia en las listas, característica que si aprovechan los demás métodos. Por otro lado, es claro que este método sí permite aprovechar la complementariedad entre las listas de respuestas, propiedad que los otros métodos sopesan de manera más indirecta.

Otra observación relevante es respecto al comportamiento del método RSV. Este método reordena las respuestas en función de la puntuación calculada en los pasos anteriores. Como puede advertirse el método RSV no permite extraer más respuestas correctas de las que se obtienen con el ejercicio monolingüe, de ahí la importancia de tratar de hacer comparables las calificaciones de las repuestas de las diferentes listas, como lo demuestran los resultados de los métodos *CombSum* y *CombMNZ*.

Por último, respecto a los resultados alcanzados por *CombSum* y *CombMNZ* se nota un mejor comportamiento que el experimento monolingüe con precisión a 3 y 5 posiciones. Una probable explicación del porqué no lo mejoran en la precisión a la 1ª posición sería los problemas durante la traducción automática de las respuestas de italiano y francés al español.

Hay que recordar que las listas de respuestas son de unas cuantas palabras y en muchas ocasiones entidades nombradas, situación que complica su correcta traducción.

<i>Método</i>	<i>1a.Pos.</i>	<i>3a.Pos</i>	<i>5a.Pos.</i>
RSV	0.49	0.67	0.73
RoundRobin	0.51	0.77	0.84
CombSum	0.48	0.77	0.83
CombMNZ	0.52	0.73	0.80

Tabla 3. Precisión con los diferentes métodos de fusión al considerar únicamente las preguntas con respuesta en más de una colección.

Como se mencionó en la sección anterior, se identificó el subconjunto de preguntas cuyas respuestas podían encontrarse en más de una colección. La tabla 3 muestra los resultados alcanzados con los métodos de fusión sobre este subconjunto de preguntas. Como era de esperarse se tienen mejores precisiones que al tomar todas las preguntas. Los métodos *CombSum* y *CombMNZ* mejoran su comportamiento, ya que estos métodos aprovechan la repetición y complementariedad de las listas de respuestas. Sin embargo, es nuevamente notorio el comportamiento del método de *RoundRobin*.

5 Conclusiones

En este trabajo se expusieron los principales desafíos en la construcción de sistemas de BR Multilingües y se abundó sobre el problema de fusión de listas de respuestas en diferentes idiomas. Finalmente, se evaluaron las técnicas de fusión usadas comúnmente en los sistemas de Recuperación de Información Multilingües.

Con estos experimentos se demostró la factibilidad y utilidad de los sistemas de BR multilingües, ya que se obtienen mejores precisiones a 3 y 5 posiciones. Sin embargo, es claro que los métodos de fusión tradicionales no colocan en la primera posición la respuesta correcta. Creemos que esto se debe fundamentalmente al error introducido por una mala traducción de las respuestas y, en menor medida, a la cantidad de respuestas consideradas en el proceso de fusión. De ahí la necesidad de proponer un nuevo método de fusión específico a esta problemática. Este método deberá: (i) considerar la complementariedad de las listas, (ii) disminuir

el impacto de una mala traducción (usando, por ejemplo, varios traductores), e (iii) integrar en el esquema de fusión la calificación otorgada por los sistemas de BR monolingües.

Agradecimientos

Este trabajo fue parcialmente financiado por el CONACYT bajo el proyecto 43990 y con la beca 184663. También agradecemos al CLEF por los recursos proporcionados.

Bibliografía

- Aceves-Pérez R., Montes-y-Gómez M, Villaseñor-Pineda L, 2007. Enhancing Cross-Language Question Answering by Combining Multiple Question Translations. *Lecture Notes in Computer Science*, Vol. 4394, Springer 2007.
- Chu-Carroll J., Czuba K., Prager A.J., Ittycheriah A. 2003. In question answering, two heads are better than one. *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1*. Edmonton, Canada. 2003.
- Echihabi A., Oard D., Marcu D. y Hermjakob U., 2003. Cross-language question answering at the USC Information Sciences Institute. *Lecture Notes in Computer Science* Vol. 3237, Springer 2004.
- García-Cumbreras M. A., Ureña-López L. A., Martínez-Santiago F., Perea-Ortega J. M. 2006. BRUJA System. The University of Jaén at the Spanish Task of CLEFQA 2006. In *Working Notes of CLEF 2006*, Alicante, España. 2006.
- Jijkoun V., Mishne G., Rijke M., Schlobach S., Ahn D., Muller K., 2004. The University of Amsterdam at QA@CLEF 2004. In *Working Notes of CLEF 2004*, Bath, UK, 2004.
- Laurent D., Séguela P., and Nègre S., 2005 Cross lingual question answering using QRISTAL for CLEF 2005. In *Working Notes CLEF 2005*. Vienna, Austria. 2005.
- Lee J., 1997. Analysis of Multiple Evidence Combination, *20th annual ACM SIGIR Conference Proceedings*, 1997.
- Magnini B., Romagnoli S., Vallin A., Herrera J., Peñas A., Peinado V., Verdejo F. and

- Rijke M., 2003. In *Working Notes of CLEF 2003*, Trondheim, Norway, 2003.
- Montes-y-Gómez, M., Villaseñor-Pineda, L., Pérez-Coutiño, M., Gómez-Soriano, J. M., Sanchis-Arnal, E. & Rosso, P., 2005. INAOE-UPV Joint Participation in CLEF 2005: Experiments in Monolingual Question Answering. In *Working Notes of CLEF 2005*, Vienna, Austria, 2005.
- Neumann G. and Sacaleanu B., 2005. DFKI's LT-lab at the CLEF 2005 multiple language question answering track. In *Working Notes CLEF 2005*, Vienna, Austria. 2005.
- Perret L., 2004. Question answering system for the French. *Lecture Notes in Computer Science*, Vol. 3491, 2005.
- Sangoi Pizzato L. A., and Molla-Aliod D., 2005. Extracting Exact Answers using a Meta Question Answering System. In *Proceedings of the Australasian Language Technology Workshop 2005*, Sydney, Australia, 2005.
- Sutcliffe R., Mulcahy M., Gabbay I., O'Gorman A., White K., Slatter D., 2005. Cross-Language French-English Question Answering using the DLT System at CLEF 2005. In *Working Notes CLEF 2005*. Vienna, Austria. 2005.
- Tanev H, Negri M., Magnini B., and Kouylekov M., 2004. The DIOGENE question answering system at CLEF-2004. In *Working Notes of CLEF 2004*, Bath UK. 2004.